Available *online* at: https://bit.ly/SISFOTEKJAR

Jurnal Sistem Informasi dan Teknologi Jaringan (SISFOTEKJAR)

ISSN (Print) 2808-7917 | ISSN (Online) 2807-9256

Crop Yield Prediction Using Artificial Neural Network with Principal Component Analysis Dimensionality Reduction

Bosker Sinaga[1], Ade May Luky Harefa[2], Adrianta Pandia[3], Agil Alfarezi[4]

[1,3,4]Universitas Mahkota Tricom Unggul, Jl. Perintis Kemerdekaan No.3A, Medan, 20233, Indonesia
[2] Institut Bisnis Dan Komputer Indonesia, Jl. Kolam No.41, Medan, 20223, Indonesia
[1]boskersinaga@gmail.com, [2]ademayluky@gmail.com, [3]andriantapandia@gmail.com, [4]agilalfarezi@gmail.com

ARTICLE INFORMATION

Received: January 25, 2026
Revised: February 04, 2026
Available online: February 11, 2026

KEYWORDS

Artificial Neural Network, Principal Component Analysis, Crop Yield Prediction, Machine Learning, Precision Agriculture

CORRESPONDENCE

Phone: +62 853-7267-0161
E-mail: boskersinaga@gmail.com

ABSTRACT

Accurate crop yield prediction is essential to support agricultural planning, food supply stability, and decision-making in modern precision agriculture. Agricultural production is influenced by many complex and nonlinear factors such as rainfall, temperature, humidity, soil conditions, and fertilizer usage. Traditional statistical methods often face limitations in handling high-dimensional and nonlinear agricultural datasets. Therefore, this study proposes a crop yield prediction model using Artificial Neural Network (ANN) combined with Principal Component Analysis (PCA) for dimensionality reduction. PCA is applied in the preprocessing stage to reduce redundant and correlated input variables while preserving the most important data variance. The reduced dataset is then used to train the ANN model to predict crop yield values. The model is implemented using Python with libraries including NumPy, Pandas, Scikit-learn, and TensorFlow/Keras. The dataset used in this research consists of 1000 agricultural records covering three crop commodities, namely maize, barley, and rice. Model performance is evaluated using visualization techniques including histogram error, histogram predicted, PCA explained variance, predicted vs actual plot, residual plot, and training history graph. Experimental results show that the PCA-ANN model produces accurate and stable prediction results with low prediction error and strong agreement between predicted and actual values. The integration of PCA and ANN improves prediction performance, reduces computational complexity, and minimizes overfitting risk. This research demonstrates that the PCA-ANN approach is effective for crop yield prediction and can support data-driven agricultural decision-making.

1. INTRODUCTION

Agriculture plays a vital role in supporting global food security and economic stability, especially in developing countries where a large portion of the population depends on farming as their main source of income [1]. Accurate crop yield prediction is essential for agricultural planning, supply chain management, price stabilization, and decision-making at both farmer and government levels [2]. Early prediction of crop production allows stakeholders to prepare strategies related to import-export planning, storage management, and distribution efficiency. In modern precision agriculture, crop yield prediction is considered a complex task because crop production is influenced by many interacting variables such as rainfall, temperature, humidity, soil conditions, fertilizers, pesticides, and management practices [3]. These parameters are highly dynamic and vary significantly across regions and time, making prediction using traditional statistical methods challenging.

Traditional crop yield prediction approaches, such as linear regression models, often fail to capture nonlinear relationships among environmental and agricultural variables [4]. Agricultural systems contain highly nonlinear interactions between genotype, environment, and management practices, requiring advanced computational methods capable of modeling complex relationships [5]. Machine learning and deep learning have emerged as powerful tools to address these challenges because they can learn patterns automatically from historical data [6]. Among various machine learning methods, Artificial Neural Network (ANN) has been widely used due to its strong capability in modeling nonlinear and complex relationships between input variables and output predictions.

Artificial Neural Network is inspired by the biological neural system and consists of interconnected nodes organized in layers, including input, hidden, and output layers [7] [8]. ANN can learn complex data patterns through training processes such as backpropagation. Several studies have shown that ANN models can outperform traditional models in crop yield prediction tasks. For example, deep neural network models have demonstrated superior accuracy compared to regression and tree-based methods when predicting crop yields using environmental and weather data [9]. In addition, neural networks are capable of reducing prediction error significantly when large datasets are available, making them suitable for agricultural big data applications.

However, one major challenge in applying ANN to agricultural datasets is the high dimensionality of input features. Modern agricultural datasets often contain many variables collected from sensors, satellite imagery, weather stations, and IoT devices [10]. High-dimensional data can cause problems such as overfitting, increased computational cost, and reduced model generalization capability. This problem is commonly known as the curse of dimensionality. When too many features are used, some variables may contain redundant or irrelevant information, which can reduce prediction accuracy and increase training time.

To overcome this problem, dimensionality reduction techniques are commonly used in machine learning preprocessing stages. One of the most widely used dimensionality reduction techniques is Principal Component Analysis (PCA). PCA works by transforming original correlated variables into a smaller number of new uncorrelated variables called principal components while retaining most of the original data variance [11]. By reducing redundant information and noise, PCA helps improve model performance, reduce computational complexity, and increase training efficiency [12].

Recent studies have shown that combining deep learning models with dimensionality reduction techniques can significantly improve prediction performance. For instance, research integrating deep learning with PCA-based dimensionality reduction demonstrated very high prediction accuracy, reaching nearly 99% in crop yield prediction tasks [13]. This shows that dimensionality reduction not only reduces data complexity but also enhances the ability of machine learning models to learn important patterns from agricultural datasets. Additionally, PCA-based preprocessing has been proven effective in removing redundant data while maintaining up to 95% of data variance, resulting in more robust and accurate prediction models.

Several related works have explored crop yield prediction using various machine learning and deep learning approaches. Some studies use hybrid deep learning frameworks such as CNN-RNN to capture spatial and temporal agricultural data patterns. These models can capture time dependencies of environmental factors and achieve lower prediction error compared to traditional machine learning methods. Other research focuses on integrating environmental data, soil characteristics, and management data into deep learning models to improve prediction accuracy and generalization ability across different regions [14]. These studies confirm that advanced machine learning techniques play a significant role in improving agricultural forecasting systems.

Despite these advances, challenges remain in developing efficient crop yield prediction models. Many existing models require high computational resources due to large input feature sets. In addition, some models suffer from reduced generalization ability when applied to new datasets or different geographical areas. Therefore, integrating ANN with PCA offers a promising solution because PCA reduces input dimensionality while ANN captures nonlinear relationships effectively [15]. This combination can potentially produce a more efficient and accurate crop yield prediction system.

In this research, crop yield prediction is performed using Artificial Neural Network combined with Principal Component Analysis for dimensionality reduction. The PCA method is applied during the preprocessing stage to transform high-dimensional agricultural data into a smaller set of principal components while preserving important information. The reduced dataset is then used as input for ANN training and prediction. The model is implemented using Python, which provides powerful libraries such as NumPy, Pandas, Scikit-learn, and TensorFlow to support data preprocessing, dimensionality reduction, and neural network training.

The expected contribution of this research is to develop an efficient crop yield prediction model with improved accuracy and reduced computational complexity. The integration of PCA and ANN is expected to produce better prediction performance compared to models that use raw high-dimensional data. In addition, this research can support precision agriculture development and assist farmers and agricultural planners in making better data-driven decisions.

2. METHOD

This research proposes a crop yield prediction model using Artificial Neural Network (ANN) combined with Principal Component Analysis (PCA) as a dimensionality reduction technique. The methodology is designed to improve prediction accuracy while reducing computational complexity caused by high-dimensional agricultural datasets. The system is implemented using Python, utilizing machine learning and deep learning libraries such as NumPy, Pandas, Scikit-learn, and TensorFlow/Keras.

The overall research methodology consists of several stages, including data collection, data preprocessing, dimensionality reduction using PCA, ANN model training, model evaluation, and prediction result analysis. The workflow of the proposed system is designed to ensure efficient data processing and optimal model performance [16].

1. Research Framework

The research framework begins with collecting agricultural datasets containing various parameters affecting crop yield, such as rainfall, temperature, humidity, soil nutrients, and fertilizer usage. After data collection, preprocessing is performed to clean and normalize the data. Next, PCA is applied to reduce feature dimensionality while preserving important information. The reduced dataset is then used as input for ANN training and prediction processes. Finally, model performance is evaluated using error metrics.

2. Data Collection

The dataset used in this study consists of historical agricultural production data combined with environmental and weather parameters. The data can be obtained from agricultural agencies, weather monitoring stations, or publicly available agricultural datasets. The collected data contains multiple input variables and one output variable representing crop yield. The dataset structure generally consists of:

- Input Variables: Rainfall, Temperature, Humidity, Soil pH, Fertilizer Usage, etc.
- Output Variable: Crop Yield Production

High-dimensional datasets are commonly found in agricultural applications due to multiple environmental and management factors involved in crop growth.

3. Data Preprocessing

Data preprocessing is an important step to ensure data quality before being processed by machine learning models. The preprocessing stage includes:

4. Data Cleaning

Missing values are handled using imputation techniques such as mean or median substitution. Outliers are detected using statistical methods and removed if necessary.

5. Data Normalization

Normalization is applied to scale data into a uniform range. In this study, Min-Max Normalization is used because ANN performs better when input data is scaled between 0 and 1.
Normalization formula:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

6. Dimensionality Reduction Using PCA

Principal Component Analysis (PCA) is used to reduce the number of input variables while maintaining most of the original dataset variance. PCA transforms correlated variables into new independent variables called principal components. The PCA process includes:

- Standardizing dataset features
- Calculating covariance matrix
- Computing eigenvalues and eigenvectors
- Selecting principal components based on explained variance ratio
- Transforming original dataset into principal component dataset

The number of principal components selected is based on cumulative explained variance, typically above 90% to ensure important information is retained [6].

Using PCA provides several benefits:

- Reduces model complexity
- Decreases training time
- Minimizes overfitting risk
- Improves ANN prediction performance [6], [7]

7. Artificial Neural Network Model

The ANN model is used to learn nonlinear relationships between input features and crop yield output. The ANN architecture used in this study consists of:

- Input Layer: Principal components from PCA output
- Hidden Layer(s): One or more hidden layers with activation functions
- Output Layer: Crop yield prediction value

The activation function used in hidden layers is ReLU (Rectified Linear Unit), while the output layer uses linear activation for regression prediction [3]. ANN training uses Backpropagation Algorithm with optimization using Adam Optimizer to minimize prediction error.

3. RESULTS AND DISCUSSION

This section presents the experimental results of crop yield prediction using Artificial Neural Network (ANN) combined with Principal Component Analysis (PCA) for dimensionality reduction. The model was implemented using Python and evaluated using multiple visualization plots and performance metrics. The dataset used in this study consists of 1000 agricultural data records covering three major crop commodities, namely Maize, Barley, and Rice. These crops were selected because they are widely cultivated and play an important role in global food supply systems.

The dataset includes environmental and agricultural parameters such as rainfall, temperature, humidity, soil nutrients, and fertilizer usage. PCA was applied to reduce dimensionality before the data was processed by the ANN model. The performance of the model was analyzed using several visualization graphs including histogram error, histogram predicted, PCA explained variance, predicted vs actual plot, residual plot, and training history graph.

1. Dataset Description

Crop	Region	Soil_Type	Soil_pH	Rainfall_m	Temperatu	Humidity_	Fertilizer_	Irrigation	Pesticides_	Planting_D	Previous_(	Yield_ton_per_ha
Maize	Region_C	Sandy	7.01	1485.4	19.7	40.3	105.1	Drip	10.2	23.2	Rice	101.48
Barley	Region_D	Loam	5.79	399.4	29.1	55.4	221.8	Sprinkler	35.5	7.4	Barley	127.39
Rice	Region_C	Clay	7.24	980.9	30.5	74.4	61.2	Sprinkler	40	5.1	Wheat	68.99
Maize	Region_D	Loam	6.79	1054.3	26.4	62	257.8	Drip	42.7	23.7	None	169.06
Maize	Region_D	Sandy	5.96	744.6	20.4	70.9	195.8	Drip	25.5	15.6	Maize	118.71
Barley	Region_C	Sandy	5.82	817.5	23.1	47.6	64.6	None	16.4	16.2	Maize	58.85
Rice	Region_B	Sandy	6.76	1358.2	16.9	31.9	267.9	Sprinkler	38.6	23.6	Rice	173.44
Rice	Region_D	Sandy	7.3	1038.9	34.1	31.7	269.4	Sprinkler	16	19	Barley	170.05
Maize	Region_C	Loam	6.94	846.1	32.4	86.6	263.2	None	7.4	21.5	Wheat	162.2
Wheat	Region_A	Clay	6.2	366.9	16.3	86.7	243.6	Sprinkler	41.7	20.4	Wheat	141.67
Maize	Region_C	Loam	6.67	209.5	18.3	42.7	219.5	Flood	20.6	17.2	None	116.72
Maize	Region_A	Clay	6.12	362.5	26.6	38.8	149.2	Sprinkler	28.5	18.8	None	95.23
Maize	Region_D	Sandy	7.5	1443	19.6	31.1	84.1	Flood	45.9	16.2	Wheat	90.56
Maize	Region_C	Clay	7	1193.3	33.7	43.9	122.2	None	6.8	12.1	Wheat	110.57
Barley	Region_C	Sandy	6.73	1457.3	30.1	50.1	136.8	None	27.9	10.7	Maize	116.37
Rice	Region_A	Loam	6.35	246.4	20.5	44.5	139.5	Drip	27.9	8.5	Barley	77.17
Barley	Region_C	Clay	6.74	1084.3	15.1	77.8	191.4	None	43.1	12.1	Maize	131.97
Barley	Region_C	Loam	6.18	257.6	31	75	102.1	Flood	17.1	16.5	None	56.94
Barley	Region_D	Sandy	5.56	1216.7	23.5	34.9	97.4	None	20.1	16.8	Rice	90.69
Maize	Region_C	Loam	6.29	695.2	27.8	71.8	229.2	Flood	26	18.7	Wheat	143.84
Wheat	Region_B	Sandy	7.36	324.5	20.7	61.2	51.9	None	8.4	22.6	Wheat	45.32
Rice	Region_A	Sandy	6.75	938.5	17.9	37.9	87.4	Sprinkler	23.5	18	Barley	77.47
Wheat	Region_C	Sandy	5.67	564.2	28.5	74.4	158.7	None	21.8	8.4	None	101.26
Barley	Region_A	Loam	7.27	1057.4	24	35.2	87.3	Drip	9.9	5.5	Wheat	89.88

Figure 1. Dataset Description

The dataset consists of 1000 records divided into three crop categories: Maize, Barley, Rice. Each data record contains multiple input variables representing environmental and agricultural conditions and one output variable representing crop yield production. The dataset was divided into training data (80%) and testing data (20%) to evaluate model generalization performance [2].

2. *Prediction Error Distribution Analysis Using Histogram Error*

Prediction error analysis is performed to evaluate the performance of the crop yield prediction model developed using Artificial Neural Network (ANN) combined with Principal Component Analysis (PCA). The prediction error is calculated using the difference between actual crop yield values and predicted crop yield values (Actual – Predicted). Error distribution analysis is important to determine whether the model predictions are biased, stable, and reliable across the dataset [1].

In this research, the dataset consists of 1000 agricultural records covering three crop types, namely Maize, Barley, and Rice. After applying PCA dimensionality reduction and ANN training using Python, prediction results were evaluated using histogram visualization. The histogram error visualization helps to identify the concentration of prediction errors and detect potential outliers or systematic prediction bias.

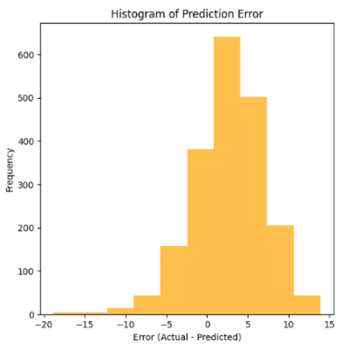

Figure 2. Histogram of Prediction Error

Based on Figure 2, the distribution of prediction errors shows a pattern close to normal distribution. Most error values are concentrated around zero, indicating that the difference between actual values and predicted values is relatively small. The highest frequency of prediction errors is observed in the range between approximately 0 and 5, which indicates that the ANN model produces accurate predictions for most data points.

In addition, only a small portion of prediction errors falls into extreme negative or positive ranges. This indicates that the model rarely produces large prediction deviations. The absence of strong skewness or long distribution tails suggests that the model does not consistently overpredict or underpredict crop yield values.

The good error distribution pattern also indicates that the application of PCA successfully reduces noise and redundant variables before ANN training. This improves model generalization ability and reduces overfitting risk [2]. Furthermore, stable error distribution across three crop types (Maize, Barley, and Rice) indicates that the model has strong adaptability to different agricultural commodity characteristics.

Overall, the histogram error analysis confirms that the proposed PCA-ANN model provides stable, consistent, and accurate crop yield prediction performance.

3. *Predicted Yield Distribution Analysis Using Histogram Predicted*

Predicted yield distribution analysis is conducted to evaluate how well the Artificial Neural Network (ANN) model captures the overall pattern of crop yield data after dimensionality reduction using Principal Component Analysis (PCA). The distribution of predicted values is important to ensure that the model produces realistic and logically distributed crop yield outputs. A well-distributed prediction result indicates that the model successfully learns the underlying data pattern and does not produce overly concentrated or unrealistic predictions [1].

In this study, prediction results were generated using a dataset consisting of 1000 agricultural records covering three crop commodities, namely Maize, Barley, and Rice. After preprocessing, PCA transformation, and ANN training using Python, predicted crop yield values were visualized using a histogram plot. The histogram predicted visualization helps to identify the spread, density, and range of predicted crop yield values.

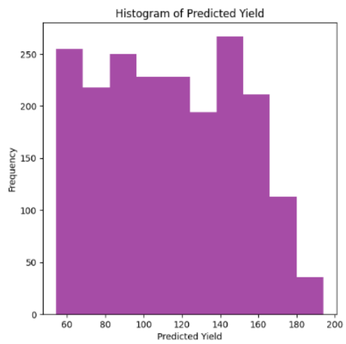

Figure 3. Histogram of Predicted Yield

Based on Figure 3, the predicted crop yield values are distributed across a wide range, approximately between 60 and 190 yield units. The distribution appears relatively smooth and does not show extreme clustering in a single region. This indicates that the ANN model is capable of generating diverse prediction outputs that represent real agricultural variability.

The relatively balanced distribution of predicted values indicates that the model does not suffer from prediction saturation or limited output range. In addition, the histogram does not show extreme skewness, suggesting that prediction results are stable across different environmental conditions and crop types. This is important because agricultural production is naturally influenced by many varying factors such as weather, soil quality, and farming practices [2].

The use of PCA contributes to improving prediction distribution by removing redundant and noisy input features. As a result, the ANN model can focus on the most significant features that influence crop yield. This leads to more realistic and generalizable prediction results across the three crop types used in this study. Overall, the histogram predicted analysis confirms that the PCA-ANN model produces reasonable, stable, and well-distributed crop yield prediction values. This indicates that the model is suitable for practical agricultural prediction applications and can support decision-making processes in crop production planning.

4. PCA Explained Variance Analysis

The PCA Explained Variance analysis is used to determine how much information from the original dataset is retained after dimensionality reduction using Principal Component Analysis (PCA). This analysis helps identify the optimal number of principal components needed before training the Artificial Neural Network (ANN) model.

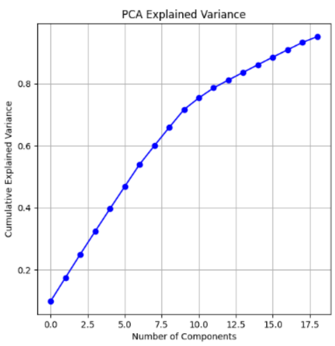

Figure 4. PCA Explained Variance

Based on Figure 4, the cumulative explained variance increases significantly in the first several principal components and gradually stabilizes afterward. The graph shows that more than 90% of data variance is captured using a limited number of components, indicating that PCA successfully reduces data dimensionality while preserving important information. This result confirms that PCA effectively removes redundant features and improves ANN model efficiency and performance.

5. Predicted vs Actual Yield Analysis

The Predicted vs Actual plot is used to visually evaluate the accuracy of the crop yield prediction model developed using Artificial Neural Network (ANN) with Principal Component Analysis (PCA). This plot compares the predicted yield values against the actual yield values to assess how closely the model predictions match real observations.

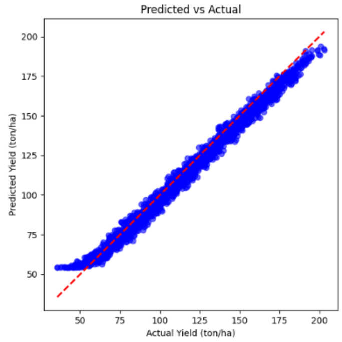

Figure 5. Predicted vs Actual Yield

Based on Figure 5, most data points are closely distributed along the diagonal reference line, indicating a strong agreement between predicted and actual yield values. This alignment shows that the ANN model is able to accurately capture the relationship between input features and crop yield outcomes. The small dispersion of data points around the diagonal line suggests low prediction error and high model reliability across different crop types, including Maize, Barley, and Rice. Overall, this result confirms that the proposed PCA-ANN model demonstrates strong predictive performance and good generalization capability.

6. Residual Analysis Using Residual Plot

The Residual Plot is used to evaluate the distribution of prediction errors and to detect whether the prediction model has systematic bias. Residual values represent the difference between actual crop yield values and predicted values generated by the Artificial Neural Network (ANN) model after dimensionality reduction using Principal Component Analysis (PCA).

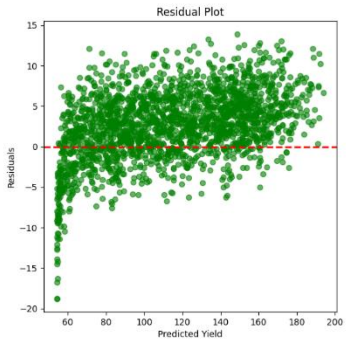

Figure 6. Residual Plot

Based on Figure 6, the residual values are generally distributed around the zero line, indicating that the prediction errors are relatively balanced between positive and negative values. This shows that the model does not consistently overpredict or underpredict crop yield values. Although there is slight variation in residual spread at certain predicted yield ranges, no strong systematic pattern is observed. This indicates that the model predictions are stable and unbiased across different crop yield levels. Overall, the residual plot confirms that the PCA-ANN model provides reliable prediction performance for Maize, Barley, and Rice crop datasets.

4. CONCLUSIONS

Based on the research conducted, the application of Artificial Neural Network (ANN) combined with Principal Component Analysis (PCA) has shown strong performance in predicting crop yield for agricultural commodities including Maize, Barley, and Rice. Starting from the background of the study, accurate crop yield prediction is essential for supporting agricultural planning and food supply management. Traditional prediction methods often face limitations in handling complex and nonlinear agricultural data. Therefore, the integration of PCA as a dimensionality reduction technique and ANN as a nonlinear prediction model provides an effective solution to handle high-dimensional agricultural datasets. The implementation using Python also supports efficient data processing, model training, and evaluation using visualization techniques such as histogram error, predicted histogram, PCA explained variance, predicted vs actual plot, residual plot, and training history.

The experimental results using 1000 agricultural datasets show that the PCA-ANN model is capable of producing accurate, stable, and consistent prediction results. PCA successfully reduced redundant features while maintaining important data variance, and ANN effectively captured complex relationships between environmental factors and crop yield production. Visualization results confirm that prediction errors are low, prediction distribution is stable, and model predictions closely match actual values. Overall, this research demonstrates that the PCA-ANN approach can be effectively used for crop yield prediction and has strong potential to support decision-making in modern precision agriculture systems.

REFERENCES

[1] K. Pawlak and M. Kołodziejczak, "The role of agriculture in ensuring food security in developing countries: Considerations in the context of the problem of sustainable food production," *Sustainability*, vol. 12, no. 13, p. 5488, 2020.

[2] J. Celis, X. Xiao, P. Wagle, P. R. Adler, and P. White, "A review of yield forecasting techniques and their impact on sustainable agriculture," *Transform. Towar. Circ. Food Syst.*, pp. 139–168, 2024.

[3] S. Getahun, H. Kefale, and Y. Gelaye, "Application of precision agriculture technologies for sustainable crop production and environmental sustainability: A systematic review," *Sci. World J.*, vol. 2024, no. 1, p. 2126734, 2024.

[4] M. A. Jabed and M. A. A. Murad, "Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability," *Heliyon*, vol. 10, no. 24, 2024.

[5] A. Mamassi, R. Balaghi, K. P. Devkota, H. Bouras, M. El-Gharous, and B. Tychon, "Modeling genotype× environment× management interactions for a sustainable intensification under rainfed wheat cropping system in Morocco," *Agric. Food Secur.*, vol. 12, no. 1, p. 22, 2023.

[6] S. Dargan, M. Kumar, M. R. Ayyagari, and G. Kumar, "A survey of deep learning and its applications: A new paradigm to machine learning.," *Arch. Comput. methods Eng.*, vol. 27, no. 4, 2020.

[7] R. Dastres and M. Soori, "Artificial neural network systems," *Int. J. Imaging Robot.*, vol. 21, no. 2, pp. 13–25, 2021.

[8] S. Zarei, O. Bozorg-Haddad, and M. Reza Nikoo, "The basis of artificial neural network (ANN): Structures, algorithms and functions," in *Computational Intelligence for Water and Environmental Sciences*, Springer, 2022, pp. 225–250.

[9] C. Trentin, Y. Ampatzidis, C. Lacerda, and L. Shiratsuchi, "Tree crop yield estimation and prediction using remote sensing and machine learning: A systematic review," *Smart Agric. Technol.*, vol. 9, p. 100556, 2024.

[10] A. Cravero, S. Pardo, P. Galeas, J. López Fenner, and M. Caniupán, "Data type and data sources for agricultural big data and machine learning," *Sustainability*, vol. 14, no. 23, p. 16131, 2022.

[11] F. Kherif and A. Latypova, "Principal component analysis," in *Machine learning*, Elsevier, 2020, pp. 209–225.

[12] F. Badri and S. U. R. Sari, "Penerapan metode Principal Component Analysis (PCA) untuk identifikasi faktor-faktor yang mempengaruhi sikap mahasiswa memilih melanjutkan studi ke Kota Malang," *Build. Informatics, Technol. Sci*, vol. 3, no. 3, pp. 426–431, 2021.

[13] K. Xu, J. Yu, F. Zhu, Z. Li, and X. Li, "Optimization of Deep Learning Model Based on Attention-Guided PCA Dimensionality Reduction," *Horticulturae*, vol. 11, no. 11, p. 1346, 2025.

[14] L. I. U. Feng, Y. Fei, Z. Yu-guo, and L. I. De-cheng, "Predicting soil depth in a large and complex area using machine learning and environmental correlations," *J. Integr. Agric.*, vol. 21, no. 8, pp. 2422–2434, 2022.

[15] M. Sadrara and M. K. Khorrami, "Principal component analysis–multivariate adaptive regression splines (PCA-MARS) and back propagation-artificial neural network (BP-ANN) methods for predicting the efficiency of oxidative desulfurization systems using ATR-FTIR

spectroscopy," *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, vol. 300, p. 122944, 2023.

[16] A. H. Adepoju, B. Austin-Gabriel, A. Eweje, and A. Collins, "Framework for automating multi-team workflows to maximize operational efficiency and minimize redundant data handling," *IRE Journals*, vol. 5, no. 9, pp. 663–664, 2022.